AD-A063 317 HYDROTRONICS FALLS CHURCH VA F/G 9/3
TIME CONSTANTS AND LEARNING CURVES OF LMS ADAPTIVE FILTERS.(U)
NOV 78 M SHENSA N00123-76-C-1937
UNCLASSIFIED NOSC-TR-312 NL

OF
AD
A063317

END
DATE
FILMED
3-79
DDC

NOSC

NOSC TR 312

Technical Report 312

TIME CONSTANTS AND LEARNING CURVES OF LMS ADAPTIVE FILTERS

M Shensa
(Hydrotronics, Incorporated)

1 November 1978

Final Report: October 1977 - June 1978

Prepared for
Naval Electronic Systems Command

Approved for public release; distribution unlimited.

NAVAL OCEAN SYSTEMS CENTER
SAN DIEGO, CALIFORNIA 92152

NAVAL OCEAN SYSTEMS CENTER, SAN DIEGO, CA 92152

AN ACTIVITY OF THE NAVAL MATERIAL COMMAND

RR GAVAZZI, CAPT, USN
Commander

HL BLOOD
Technical Director

ADMINISTRATIVE INFORMATION

This report was prepared by Hydrotronics, Incorporated, under Contract Number N00123-76-C-1937. Work for this report was performed from October 1977 through June 1978. Jim Zeidler, Head, Analysis Branch of the Fleet Engineering Department acted as Contract Monitor. The work presented in this report was part of NOSC Project Number 62711N-F11101-XF-111-011-00.

Released by
R. H. Hearn, Head
Electronics Division

Under authority of
D. A. Kunz, Head
Fleet Engineering Department

ACKNOWLEDGEMENTS

The author wishes to thank Jim Zeidler for his comments as well as for initiating this study, and Ed Satorius for many stimulating discussions.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR 312
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): Time Constants and Learning Curves of LMS Adaptive Filters
5. TYPE OF REPORT & PERIOD COVERED: Final Report, Oct 1977 – Jun 1978
6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): M. Shensa
8. CONTRACT OR GRANT NUMBER(s): N00123-76-C-1937

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Ocean Systems Center, San Diego, CA 92152
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62711N-F11101-XF-111-011-00

11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Electronic Systems Command, Washington, D.C. 20360
12. REPORT DATE: 1 November 1978
13. NUMBER OF PAGES: 52

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

12 54 p. 18 NOSC 19 TR-312

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

16 F11101 17 XF111011 00

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper treats the convergence of adaptive LMS filters and, in particular, the adaptive line enhancer (ALE). The learning curves of such a filter are a sum of exponentially decaying modes with time constants given by the eigenvalues of the input correlation matrix and the relative initial magnitudes given by the projections of the filter on the eigenvectors. It is shown that, for large filter lengths, a simple correspondence may be set up between the discrete and continuous cases. Indexed by frequency, the eigenvalues of the correlation matrix correspond to → next page

388 907

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

the magnitude of the power spectrum, and the projections onto the eigenvectors to the filter transfer function. A detailed analysis is carried out for single pole spectra and evaluated through a computer simulation. In general, the techniques developed provide a physical context, i.e., the signal spectrum, in which to evaluate convergence. Thus, it is possible, with varying degrees of accuracy depending on knowledge of the input spectrum, to predict the convergence behavior of the system in general.

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

This paper treats the convergence of adaptive LMS filters and, in particular, the adaptive line enhancer (ALE). The learning curves of such a filter are a sum of exponentially decaying modes with time constants given by the eigenvalues of the input correlation matrix and the relative initial magnitudes given by the projections of the filter on the eigenvectors. It is shown that, for large filter lengths, a simple correspondence may be set up between the discrete and continuous cases. Indexed by frequency, the eigenvalues of the correlation matrix correspond to the magnitude of the power spectrum, and the projections onto the eigenvectors to the filter transfer function. A detailed analysis is carried out for single pole spectra and evaluated through a computer simulation. In general, the techniques developed provide a physical context, i.e., the signal spectrum, in which to evaluate convergence. Thus, it is possible, with varying degrees of accuracy depending on knowledge of the input spectrum, to predict the convergence behavior of the system in general.

I. INTRODUCTION

This study investigates the convergence properties of LMS adaptive filters for a wide class of input spectra, thus extending some results of [1] and [2]. Utilization of the asymptotic properties of Toeplitz matrices permits the development of a fairly general theory. More specific conclusions are reached for the case of the adaptive line enhancer (ALE), which is used both as an introductory example and a concrete application. In particular, a detailed analysis is carried out for single pole spectra (i.e., a narrowband signal in white noise). Several simplified derivations of previous results ([2] and [3]) are included in order that the discussion may remain self-contained.

We begin with a brief description of some basic equations (Section II). This is followed by a treatment of sinusoids in white noise (Section III), which is intended to provide some intuition for the theory of Section IV. In Section IV, it is shown that for large filter lengths, the learning curves of the discrete LMS filter may be approximated by the continuous case. In particular, a great deal may be said about the convergence process without extensive calculations, simply from some knowledge of the input spectrum. In Section V, the theory is applied to single-pole spectra and evaluated through a computer simulation.

II. BACKGROUND

The convergence time for a linear time invariant system depends, in general, on the eigenvalues of the system and its initial state. When adjustable parameters are involved, a coarse indicator of the optimal convergence rate to be obtained is the conditioning number; that is, the ratio of the largest to the smallest eigenvalue. However, inasmuch as suitably chosen initial conditions can arbitrarily prolong convergence, any further statement requires some limitation on the initial state, usually taken to be zero. This is not so restrictive as may at first appear.

More precisely, consider the implementation of the LMS adaptive filter pictured in Figure 1. For a stationary input vector X(k), the recursion equations for the mean of the linear prediction filter, W(k), are given approximately (W(k) and X(k) are assumed independent for large k - cf [1], [4]) by [1]

$$\tilde{W}(k+1) = A\tilde{W}(k) + 2\mu P \tag{1}$$

where $\tilde{W}(k) = EW(k)$ is the expected value of the L dimensional weight vector W(k) at time k, and

$$\begin{aligned} A &= I - 2\mu R \\ &= I - 2\mu E[X^T X] \\ P &= E(dX) \ , \end{aligned} \tag{2}$$

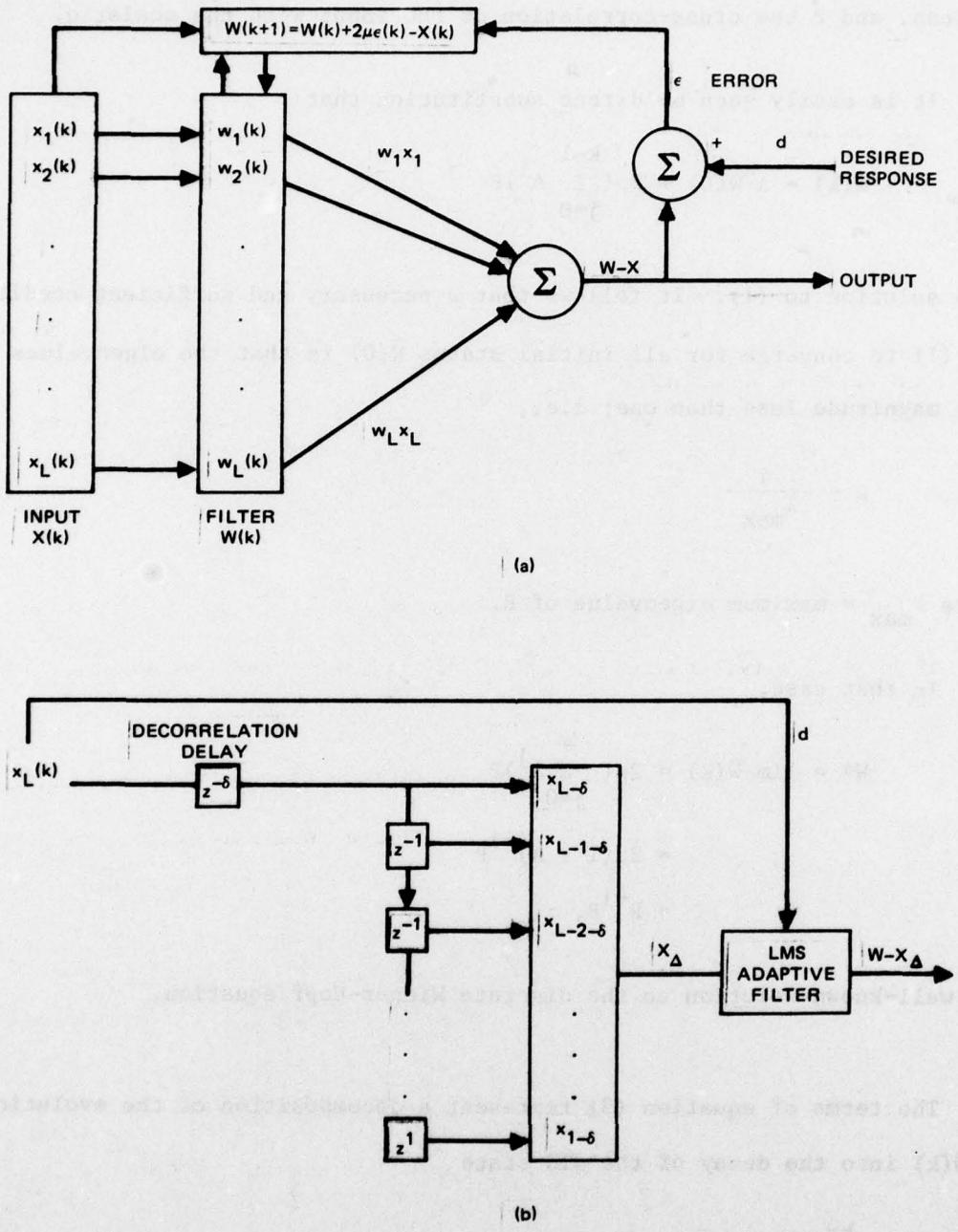

Figure 1. Block diagrams of (a) the LMS Adaptive Filter and (b) the Adaptive Line Enhancer (ALE).

with μ a feedback parameter, R the autocorrelation matrix of the input process, and P the cross-correlation of the input with the scalar d.

It is easily seen by direct substitution that

$$\tilde{W}(k) = A^k\tilde{W}(0) + 2\mu\left(\sum_{j=0}^{k-1} A^j\right)P \tag{3}$$

is a solution to (1). It follows that a necessary and sufficient condition for (1) to converge for all initial states $\tilde{W}(0)$ is that the eigenvalues of A have magnitude less than one; i.e.,

$$\mu < \frac{1}{\lambda_{max}} \tag{4}$$

where λ_{max} = maximum eigenvalue of R.

In that case,

$$\begin{aligned} W^* = \lim \tilde{W}(k) &= 2\mu\left(\sum_{j=0}^{\infty} A^j\right)P \\ &= 2\mu(I - A)^{-1}P \\ &= R^{-1}P, \end{aligned} \tag{5}$$

the well-known solution to the discrete Wiener-Hopf equation.

The terms of equation (3) represent a decomposition of the evolution of $\tilde{W}(k)$ into the decay of the old state

$$A^k\tilde{W}(0) \tag{6}$$

and the growth of the new state

$$2\mu\left(\sum_{j=0}^{k-1} A^j\right)P. \tag{7}$$

If the new state is substantially different from the old state (i.e., if its presence is easily distinguished), its detection will depend essentially on expression (7). This, and considerations of tractability, lead us to treat only the "growth rate"; i.e., we set $\tilde{W}(0)=0$. Equations (3) and (5) yield

$$\begin{aligned} W^* - \tilde{W}(k) &= 2\mu\left(\sum_{j=k}^{\infty} A^j P\right) \\ &= A^k W^* \end{aligned} \tag{8}$$

or

$$||W^* - \tilde{W}(k)||^2 = W^{*T}(A^j)^T A^j W^* \,. \tag{9}$$

The symbol T stands for conjugate transpose and $||\ ||$ indicates the vector norm. Let λ_ν, $\nu = 1, \ldots, L$ be the eigenvalues and E^ν the corresponding eigenvectors of R. Then (9) may be written

$$||W^* - \tilde{W}(k)||^2 = \sum_{\nu=1}^{L} (1-2\mu\lambda_\nu)^{2k} \, |E^\nu \cdot W^*|^2 \,. \tag{10}$$

Note that the "$\cdot$" represents the scalar product between two vectors. Equation (10) describes the convergence of $\tilde{W}(k)$ to W^*.

A similar expression may be derived for the convergence of the mean square error $\xi(k) = E(\varepsilon^2(k))$ (see Figure 1a) to its limiting value ξ^*. It is easily shown [1] that, in the absence of gradient noise (i.e., $L\mu$ small),

$$\xi(k) - \xi^* = (W^* - \tilde{W}(k))^T R(W^* - \tilde{W}(k))$$

$$= \sum_{\nu=1}^{L} (1-2\mu\lambda_\nu)^{2k} \lambda\nu |E^\nu \cdot W^*|^2 \quad . \tag{11}$$

The curves (10) and (11) as a function of k will be termed the learning curves of the filter and output respectively. Each of the L terms relaxes geometrically with a constant logarithmic slope of $\log(1-2\mu\lambda_\nu)^2$. Thus, each term falls to e^{-1} of its original value in a time given by

$$\tau_\nu = - \frac{1}{(\log(1-2\mu\lambda_\nu)^2}$$

$$\sim \frac{1}{4\mu\lambda_\nu} \text{ for } 2\mu\lambda_\nu << 1 \quad . \tag{12}$$

Inequality (4) imposes a lower bound on the largest time constant

$$\tau_{max} = \frac{1}{4\mu\lambda_{min}}$$

$$\geq \frac{\lambda_{max}}{4\lambda_{min}}$$

$$= \frac{1}{4} \text{ (conditioning number)} \quad . \tag{13}$$

Generally, τ_{max} is a very conservative estimate for the behavior of the learning curves. In practice, only those eigenvalues for which the projection of the final state on the corresponding eigenvector ($E^{\nu} \cdot W^*$ for the filter or $\lambda_{\nu} E^{\nu} \cdot W^*$ for the output) is large will exert a significant influence on convergence. This aspect will be examined in detail in succeeding sections.

III. SINUSOIDS IN WHITE NOISE (ALE)

In this section we examine the previous theory for the case of M complex sinusoids in white noise. It is shown that the L-dimensional vector space of the input may be decomposed into an M-dimensional subspace spanned by the signal component and its orthogonal complement. This permits a simplified calculation of the eigenvalues of the correlation matrix (equation 26). Then, by restricting ourselves to the ALE, we find that W* lies in the signal subspace. It follows that all eigenvectors outside that space are orthogonal to W* and, hence, their eigenvalues do not enter into the learning curves of equations (10) and (11).

One Sinusoid

Consider the case of a single sine wave in white noise $x(\ell) = \sqrt{2\sigma_s^2}\cos(\omega\ell+\theta) + \sigma_0 N(\ell)$, where θ is a uniformly distributed random variable, and $N(\ell)$ is zero-mean white noise with unit variance. Define the L-dimensional vector a by

$$a_\ell = \sigma_s e^{i\omega\ell} \quad \ell = 1, \ldots, L \ . \tag{14}$$

Then the autocorrelation matrix of x is given by

$$R = R_s + \sigma_o^2 I$$

$$= \frac{1}{2} (\phi + \overline{\phi}) + \sigma_o^2 I \tag{15}$$

where

$$\phi_{\ell k} = a_\ell \overline{a}_m = \sigma_s^2 e^{i\omega(\ell-k)} \quad . \tag{16}$$

Let us examine the effect of R on the two-dimensional subspace spanned by a and $\overline{a}$:

$$(\phi a)_\ell = \sum_{k=1}^{L} a_\ell \overline{a}_k a_k = L\sigma_s^2 a_\ell$$

$$\overline{\phi a} = L\sigma_s^2 \overline{a}$$

$$(\phi \overline{a})_\ell = \Sigma a_\ell \overline{a}_k \overline{a}_k = (\Sigma \overline{a}_k^2) a_\ell$$

$$\overline{\phi} a = (\Sigma a_k^2) \overline{a} \tag{17}$$

Thus, the subspace is invariant (is mapped into itself) under R_s, and a matrix representation of R_s in that subspace is

$$R_s \Big|_{(a,\overline{a})} = \frac{1}{2} \begin{pmatrix} L\sigma_s^2 & \Sigma \overline{a}_k^2 \\ \Sigma a_k^2 & L\sigma_s^2 \end{pmatrix} \quad . \tag{18}$$

Note that the eigenvalues of the matrix (18) will also be eigenvalues of R_s. The characteristic equation is quadratic with roots

$$\lambda_{\pm} = \frac{\sigma_s^2}{2} (L \pm |Z|)$$

$$\text{where } |Z| = |\Sigma a_m^2| = \left|\frac{\sin\omega L}{\sin\omega}\right| . \tag{19}$$

Since the columns of R_s are linear combinations of a and $\bar{a}$, its rank is at most two, and the other L-2 eigenvalues must be zero. Hence L-2 eigenvalues of R are σ_o^2 and the remaining two are $(\lambda_{\pm}) + \sigma_o^2$. Furthermore, it will be shown in the next subsection that, for the ALE of Figure 1b, W* lies in the subspace spanned by a and $\bar{a}$; and, thus, that the "growth" curves only depend on the eigenvalues $\lambda_{\pm} + \sigma_o^2$. We also note that, for ω bounded away from 0 and π, $\lambda_{\pm} \sim \sigma_s^2 L/2$, consequently, the learning curves exhibit a single time constant,

$$\tau \sim \frac{1}{4\mu(\sigma_s^2 \frac{L}{2} + \sigma_o^2)} . \tag{20}$$

This is to be contrasted with the estimate of equation (13)

$$\tau_{max} = \frac{1}{4\mu\lambda_{min}}$$

$$= \frac{1}{4\mu} \frac{1}{\sigma_o^2}$$

$$= \tau\left(\frac{L}{2} \frac{\sigma_s^2}{\sigma_o^2} + 1\right) \tag{21}$$

which, although much larger than τ except at very small signal-to-noise ratios, will not be observed when the initial state $\tilde{W}(0)$ is zero (the mode corresponding to τ_{max} is orthogonal to $\tilde{W}(0)$).

M Sinusoids

It is easy to generalize the foregoing technique to the case of M sinusoids. Define M vectors a^s, each of dimension L by

$$(a^s)_\ell = \sigma_s e^{i\omega_s \ell} \qquad \begin{matrix} s = L, \ldots, M \\ \ell = 1, \ldots, L \end{matrix} \tag{22}$$

where ω_s may be negative in order to include negative frequency components. Define the LxL matrix

$$\Phi_{\ell k} = \sum_{s=1}^{m} a_\ell^s \overline{a_k^s}$$

$$= \sum_s \sigma_s^2 e^{i\omega_s(\ell-k)} . \tag{23}$$

We now look for an eigenvector of the form $E = \sum_s \gamma_s a_s$ with eigenvalue λ

$$(\Phi E)_\ell = \sum_{r,s,k} \gamma_s a_\ell^{r-r} a_k^r a_k^s$$

$$= \lambda E_\ell$$

$$= \lambda \Sigma_r \gamma_r a_\ell^r . \tag{24}$$

Then, since the vectors a^s are linearly independent, equation (24) implies

$$\sum_s B_{rs}\gamma_s = \lambda\gamma_r. \tag{26}$$

Note that the M-dimensional subspace spanned by the vectors a^s is invariant under ϕ, that ϕ is represented by the matrix B with respect to that basis, and that B has rank M. Since ϕ also has rank M, its remaining L-M eigenvalues are zero.

Also, for the ALE of Figure 1b,

$$d(\ell) = \sum_s \sigma_s e^{i(\omega_s \ell + \theta_s)} + \sigma_o N(\ell)$$

$$x(\ell) = d(\ell) + \sigma_o N(\ell)$$

and

$$P(\ell) = \sum_s \sigma_s e^{l\omega_s(\ell+\delta-1)}, \tag{27}$$

a linear combination of the a^s (see [3] for details). Thus, $W^* = R^{-1}P = (I + B)^{-1}P$ is also contained in the subspace of the a^s. Furthermore, since ϕ is Hermition, those eigenvectors of ϕ not belonging to the matrix B are orthogonal to the above subspace and hence to W^*. We conclude that at most, M eigenvalues are pertinent to the learning curves, and they are solutions to equation (26).

Example 1

Let $a_\ell^1 = \frac{\sigma_s}{\sqrt{2}} a^{i\omega\ell}$; $a_\ell^2 = \frac{\sigma_s}{\sqrt{2}} e^{-i\omega\ell}$. Then $\phi_{\ell k} = \sigma_s^2 \cos\omega(\ell - k)$ and B is the matrix given in (18).

Example 2

$$a_\ell^1 = \frac{\sigma_1}{\sqrt{2}} e^{i\omega\ell}; \quad a_\ell^2 = \frac{\sigma_2}{\sqrt{2}} e^{i\omega\ell}; \quad a_\ell^3 = \bar{a}_\ell^1; \quad a_\ell^4 = \bar{a}_\ell^2$$

Then

$$\phi = \sigma_1 \cos\omega_1(\ell-k) + \sigma_2 \cos\omega_2(\ell-k) .$$

Let $\Delta = \omega_1 - \omega_2$ and $\omega = \omega_1 + \omega_2$. Then B is given by

$$B = \frac{1}{2} \left(\begin{array}{cc|cc} L\sigma_1^2 & \sigma_1\sigma_2 \Sigma e^{i\Delta k} & \sigma_1^2 \Sigma e^{2i\omega_1 k} & \sigma_1\sigma_2 \Sigma e^{i\omega k} \\ \sigma_1\sigma_2 \Sigma e^{-i\Delta k} & L\sigma_2^2 & \sigma_1\sigma_2 \Sigma e^{i\omega k} & \sigma_2^2 \Sigma e^{2i\omega_2 k} \\ \hline \sigma_1^2 \Sigma e^{-2i\omega_1 k} & \sigma_1\sigma_2 \Sigma e^{-i\omega k} & L\sigma_1^2 & \sigma_1\sigma_2 \Sigma e^{i\Delta k} \\ \sigma_1\sigma_2 \Sigma e^{-i\omega k} & \sigma_2^2 \Sigma e^{-2i\omega_2 k} & \sigma_1\sigma_2 \Sigma e^{i\Delta k} & L\sigma_2^2 \end{array} \right) \qquad (28)$$

We note that $\left| \sum_{k=1}^{L} e^{-ixk} \right| = \left| \frac{\sin L\frac{x}{2}}{\sin \frac{x}{2}} \right|$ which is much less than L for large L and $\frac{2}{L} < x < 2\pi - 2/L$. Thus, provided

$$\frac{2}{L} < \omega_i < \pi - \frac{1}{L} , \tag{29}$$

the off-diagonal matrices (as indicated by the dotted lines) of the entire matrix (28) may be neglected (i.e., the positive and negative frequency components uncouple). The remaining 2x2 matrices are identical and have eigenvalues

$$\lambda_{\pm} = \frac{L}{2} \frac{\sigma_1^2 + \sigma_2^2}{2} \pm \frac{1}{4} \sqrt{L^2(\sigma_1^2 - \sigma_2^2)^2 + 4|z|^2\sigma_1^2\sigma_2^2}$$

$$\text{where } |z|^2 = \left(\frac{\sin L \frac{\Delta}{2}}{\sin \frac{\Delta}{2}}\right)^2 . \tag{30}$$

For $\Delta L >> 1$; i.e., a large separation of the sinusoids compared to the filter resolution,

$$\lambda_+ = \frac{L}{2} \sigma_1^2 \qquad \Delta L >> 1$$

$$\lambda_- = \frac{L}{2} \sigma_2^2 . \tag{31}$$

On the other hand, when $\Delta L << 1$

$$\lambda_+ \sim \frac{L}{2} (\sigma_1^2 + \sigma_2^2) \qquad \Delta L << 1$$

$$\lambda_- \sim \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \frac{1}{2} \frac{L^2 - |z|^2}{L} . \tag{32}$$

and, if $\sigma_1 = \sigma_2$, λ_- reduces to

$$\lambda_- \sim \frac{\sigma_1^2}{2} \frac{L^3\Delta^2}{24} \qquad \Delta L << 1$$

$$\sigma_1^2 = \sigma_2^2 \quad . \tag{33}$$

Note that (33) could have been obtained from (19) by translating $\frac{\omega_1+\omega_2}{2}$ to zero and substituting $\frac{\Delta}{2} = \frac{\omega_1-\omega_2}{2}$ for the single frequency ω.

For two equal amplitude sinusoids close in frequency, expressions (32) and (33) are valid. In that case, even in the signal subspace, there may be a large disparity in the eigenvalues of $R = \phi + \sigma_o^2 I$:

$$\lambda_1 = \sigma_o^2(1 + qL)$$

$$\lambda_2 = \sigma_o^2(1 - q)\frac{L^3\Delta^2}{48}) \tag{34}$$

where

$$q = \frac{\sigma_s^2}{2\sigma_o^2} \tag{35}$$

is the signal-to-noise ratio. However, λ_1 or λ_2 will influence the learning curves only if the corresponding scalar product, $E_1 \cdot W^*$ or $E_2 \cdot W^*$ is relatively large (equations (10) and (11)).

The projections of the eigenvectors on W* are computed, up to a constant factor, in Appendix A with the result

$$|E^1 \cdot W^*|^2 \sim (2 + \frac{qL^3\Delta^2}{3})^2$$

$$|E^2 \cdot W^*|^2 \sim q^2\Delta^2L^4 \ . \tag{36}$$

For large q or relatively large L (note that $L\Delta << 1$ and $L^4\Delta^2 >> 1$ can be satisfied simultaneously), the second mode will have a significant effect on the weight learning curve. Its time constant

$$\tau_2 \sim \frac{1}{\sigma_o^2(1 + q\frac{L^3\Delta^2}{48})}$$

$$= \frac{(1 + qL)}{(1 + q\frac{L^3\Delta^2}{48})}\tau_1$$

$$\sim \frac{1}{\frac{1}{qL} + \frac{L^2\Delta^2}{48}}\tau_1 \tag{37}$$

is in general much longer than τ_1.

To study the output learning curve, we examine the relative magnitudes of

$$\lambda_1|E^1 \cdot W^*|^2 \sim \sigma_o^2(1 + qL)(2 + \frac{qL^3\Delta^2}{3})^2$$

$$\lambda_2|E^2 \cdot W^*|^2 \sim \sigma_o^2(1 + q\frac{L^3\Delta^2}{48})\ q^2\Delta^2L^4$$

$$= \sigma_o^2 qL(1 + q\frac{L^3\Delta^2}{48})\ qL^3\Delta^2 \ . \tag{38}$$

From (38) it is seen that the second mode is relevant only when $q\Delta^2 L^3 \geq 1$. If this condition, somewhat stronger than (36), holds, the second mode will slow the convergence of the output as well as that of the weights. These results have been confirmed experimentally [11].

IV. THEORY - GENERAL SPECTRA

For the infinite case ($L = \infty$), determining the eigenvalues and eigenvectors of the autocorrelation matrix reduces to finding its spectrum. In addition, if the transfer function of the optimal filter is known, it is not difficult to describe or approximate the learning curves. In this section, these properties are derived, and it is shown that for large L the convergence characteristics of the LMS adaptive filter are approximated by those for $L = \infty$. A more detailed treatment of the mathematics may be found in reference [5].

Let the input process $x(\ell)$ be stationary, and the z-transform [6] of its autocorrelation function given by

$$G(z) = \sum_{\ell-k=-\infty}^{\infty} E\{x(\ell)x(k)\}z^{k-\ell}$$

where E denotes expectation. Assume that $G(z)$ is continuous and has no zeroes or poles on the unit circle. (Note that the sinusoids of the previous section are a singular case.) Then the power spectrum of x may be written

$$S(\omega) = G(e^{i\omega}) \tag{40}$$

and its correlation matrix admits the representation

$$\phi_{\ell k} = \phi(\ell-k)$$

$$= E\{x(\ell)x(k)\}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) e^{i(\ell-k)\omega} d\omega \ . \tag{41}$$

Assume further, that the vector P possesses the same properties

$$P_\ell = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xd}(\omega) e^{i\ell\omega} d\omega$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} E(xd) e^{i\ell\omega} d\omega \qquad \ell = 0, \ldots, \infty \ . \tag{42}$$

For the ALE, if x is a signal process plus white noise,

$$P_\ell = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_s(\omega) e^{-i\delta\omega} e^{i\omega\ell} d\omega \ . \tag{43}$$

The finite LxL matrix of equation (2) may be written

$$R_{\ell k} = \phi\ (\ell-k) \qquad \ell, k = 0, \ldots, L-1 \ . \tag{44}$$

We introduce the infinite dimensional vector norm

$$||y||^2 = \sum_{\ell=0}^{\infty} y_\ell^2 \ . \tag{45}$$

All L-dimensional vectors may be considered as embedded in this space by setting those components with indices greater than L equal to zero.

It is shown in Appendix B that W*, the solution of (5) considered as a function of L, converges to a bounded solution f of the infinite Wiener-Hopf equation as $L \to \infty$; i.e.,

$$\sum_{k=0}^{\infty} \phi(\ell-k) f_k = P_\ell \qquad \ell = 0, \ldots, \infty \tag{46}$$

where $\lim_{L\to\infty} ||W^* - f|| = 0$.

We mention here that the matrices R and ϕ are Toeplitz [5] and, as such, possess the following properties (cf equations (B-4) and (B-9)): Let m and M be the minimum and maximum of $|S(\omega)|$ (non-zero and finite by the assumptions on G(z)); and let λ_ν, $\nu = 1, \ldots, L$ be the eigenvalues of R. Then

$$m \leq \lambda_\nu \leq M \tag{47}$$

where $$\lim_{L\to\infty} \lambda_{min} = m \qquad \lim_{L\to\infty} \lambda_{max} = M \, . \tag{48}$$

In [5] it is also shown that the matrix R is approximated (in a Hilbert-Schmidt norm) by the matrix $\hat{R}$,

$$\hat{R}_{\ell k} = \sum_{\gamma=1}^{L} S\left(\frac{2\pi\nu}{L}\right) \frac{1}{L} e^{\frac{2\pi i \nu(\ell-k)}{L}} \tag{49}$$

which has eigenvalues $S(\frac{2\pi\nu}{L})$ and normalized eigenvectors $e^{2\pi i\nu\ell/L}/\sqrt{L}$.

Thus, if E^{ν} are the eigenvectors of R, we have "approximately"

$$\lambda_{\nu} \sim S(\frac{2\pi\nu}{L}) \tag{50}$$

$$(E^{\nu})_{\ell} \sim \frac{1}{\sqrt{L}} e^{\frac{2\pi i\nu\ell}{\sqrt{L}}} \qquad \ell = 0, \ldots, L-1 . \tag{51}$$

It is, therefore, not surprising that the following expressions (derived in Appendix B) hold for the learning curves (10) and (11). Let

$$F(\omega) = \sum_{\ell=0}^{\infty} f_{\ell} e^{-i\omega\ell}, \tag{52}$$

then

$$\lim_{L\to\infty} ||W^* - \hat{W}(k)||^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} (1-2\mu S(\omega))^{2k} |F(\omega)|^2 d\omega \tag{53}$$

and

$$\lim_{L\to\infty} \xi(k)-\xi^* = \frac{1}{2\pi}\int (1-2\mu S(\omega))^{2k} S(\omega) |F(\omega)|^2 d\omega . \tag{54}$$

Note that, heuristically, from equations (50) and (51), $|E^{\nu}\cdot f|^2 \sim \frac{1}{L}\left|\sum_{\ell=0}^{L-1} f_{\ell} e^{-\frac{2\pi i\nu\ell}{L}}\right|^2 \sim \frac{1}{L}\left|F(\frac{2\pi\nu}{L})\right|^2$, and that (10) and (11) correspond to approximating sums of the integrals in (53) and (54). The factor $\frac{1}{L}$ corresponds to $\frac{d\omega}{2\pi}$. Thus, a large value of $F(\omega)$ will not be significant unless it has some bandwidth. It is the integrated power over a given spectral region that is relevant.

In summary, for large L,

the eigenvectors	are indexed by	ω,
the ω^{th} eigenvalue	corresponds to	$S(\omega)$,
and $\lvert W^{*} \cdot E_{\nu} \rvert^2$	corresponds to	$\lvert F(\omega) \rvert^2 \frac{d\omega}{2\pi}$.

The learning curves for the weight vector and output correspond to a sum of modes given by

$$\lvert F(\omega) \rvert^2 (1-2\mu S(\omega))^{2k} \frac{\Delta\omega}{2\pi}$$

and

$$\lvert F(\omega) \rvert^2 S(\omega) (1-2\mu S(\omega))^{2k} \frac{\Delta\omega}{2\pi}, \tag{55}$$

respectively, where the spectrum has been divided into regions of width $\Delta\omega$. The time constant of mode ω is thus

$$\tau(\omega) = \{4\mu S(\omega)\}^{-1} . \tag{56}$$

Finally, it follows from (47) that the conditioning number approaches

$$\frac{\lambda_{max}}{\lambda_{min}} \sim \frac{\max\limits_{\omega} S(\omega)}{\min\limits_{\omega} S(\omega)} \tag{57}$$

and the eigenvalues of R satisfy

$$\min_{\omega} S(\omega) \leq \lambda_{\nu} \leq \max_{\omega} S(\omega) . \tag{58}$$

V. SINGLE-POLE COMPLEX INPUT (ALE)

Using the previous theory, we shall now analyze in detail the learning curves of the ALE of Figure 1b for a one-pole complex input signal embedded in white noise. More precisely, assume the input autocorrelation function to be of the form

$$\phi(\ell) = \sigma_s^2 \, e^{-\alpha|\ell|} \, e^{i\omega_0 \ell} + \sigma_0^2 \, \Delta(\ell) \quad . \tag{59}$$

The z-transform of the signal alone is given by

$$G_s(z) = \sigma_s^2 \, \frac{1-|a|^2}{(z-a)(z^{-1}-\bar{a})} \tag{60}$$

where $a = e^{-\alpha+i\omega_o}$. (60)

The transform of ϕ has the form

$$G(z) = G_s(z) + \sigma_0^2$$

$$= \sigma_0^2 \, \frac{(z-b)(z-\bar{b}^{-1})}{(z-a)(z-\bar{a}^{-1})} \tag{61}$$

where b is yet to be determined.

It may be easily shown (cf [8]) by substituting (66) in (61) that

$$b = e^{-\beta}\, e^{i\omega_0} \tag{62}$$

where

$$\frac{\sigma_s^2}{\sigma_0^2} = \frac{\cosh \beta - \cosh \alpha}{\sinh \alpha} \ . \tag{63}$$

Let the signal-to-noise ratio be given by

$$q = \frac{\sigma_s^2}{\sigma_0^2} \ . \tag{64}$$

For $\alpha << 1$ (relatively narrowband) and $q\alpha << 1$, we have the approximation

$$\beta \sim \sqrt{\alpha^2 + 2\alpha q} \ . \tag{65}$$

Let us first compute the conditioning number via expression (57). From (60), on the unit circle

$$S_s(\omega) = \sigma_s^2 - \frac{1 - 2^{-2\alpha}}{e^{-2\alpha} - 2e^{-\alpha}\cos(\omega-\omega_0) + 1} \ . \tag{66}$$

Also, since

$$S(\omega) = \sigma_s^2 + S_s(\omega), \tag{67}$$

it is clear that the maximum of S occurs at $\omega = \omega_0$ and the minimum at $\omega = \omega_0 \pm \pi$.

Thus,

$$\lambda_{max} \sim \sigma_s^2 \frac{1 + e^{-\alpha}}{1 - e^{-\alpha}} + \sigma_0^2$$

$$\lambda_{min} \sim \sigma_s^2 \frac{1 - e^{-\alpha}}{1 + e^{-\alpha}} + \sigma_0^2 \quad . \tag{68}$$

For a narrowband (compared to Nyquist) signal $\alpha << 1$ and

$$\frac{\lambda_{max}}{\lambda_{min}} \sim \frac{1 + \frac{2q}{\alpha}}{1 + \frac{\alpha q}{2}} \qquad \alpha << 1 \quad . \tag{69}$$

If $q\alpha$ is also small,

$$\frac{\lambda_{max}}{\lambda_{min}} \sim 1 + \frac{2q}{\alpha} \qquad q\alpha << 1$$

$$\sim \frac{\beta^2}{\alpha^2} \tag{70}$$

It is easily shown by classical techniques ([9], [10]) that the transfer function of the optimal filter is given by

$$F(\omega) = \frac{ce^{i\omega\delta}}{1 - be^{-i\omega}}$$

$$= \frac{ce^{i\omega\delta}}{1 - e^{-\beta} e^{-i(\omega-\omega_0)}} \tag{71}$$

where c is a constant. Hence

$$|F(\omega)|^2 = \frac{c^2}{1 - 2e^{-\beta}\cos(\omega-\omega_0) + e^{-2\beta}} \quad . \tag{72}$$

Also, from (61) and (62)

$$G(z) = \sigma_0^2 \frac{\bar{b}^{-1}}{\bar{a}^{-1}} \frac{(z-b)(z^{-1}-\bar{b})}{(z-a)(z^{-1}-\bar{a})}$$

$$\Rightarrow \quad S(\omega) = \sigma_0^2 \frac{e^{\beta}}{e^{\alpha}} \frac{1 - 2e^{-\beta}\cos(\omega-\omega_0) + e^{-2\beta}}{1 - 2e^{-\alpha}\cos(\omega-\omega_0) + e^{-2\alpha}}$$

$$= \sigma_0^2 \frac{\cosh\beta - \cos(\omega-\omega_0)}{\cosh\alpha - \cos(\omega-\omega_0)} \quad . \tag{73}$$

For small β, the function (72) is very sharp and only the region $[\omega_0-\beta, \omega_0+\beta]$ will influence the integrals (53) and (54). As β increases the spectrum of F flattens out until we must include the entire interval $[-\pi, \pi]$. Let $\beta' = \beta-\alpha$. To approximate the learning curves, we divide the interval $\omega_0 \pm 2\beta'$ into three regions: $(\omega_0 - \alpha, \omega_0 + \alpha)$, $(\omega_0 - 2\beta'-\alpha, \omega_0 - \alpha)$, and $(\omega_0 + \alpha, \omega_0 + 2\beta'+\alpha)$. They are centered at ω_0, $\omega_0-\beta$, and $\omega_0+\beta$; and have widths $(\Delta\omega)_1 = 2\alpha$, $(\Delta\omega)_2 = (\Delta\omega)_3 = 2\beta'$ respectively. This is shown diagrammatically in Figure 2.

For small β and ω, $\cosh\beta \sim 1 + \frac{\beta^2}{2}$ and $\cos\omega \sim 1 - \frac{\omega^2}{2}$; and (73) may be approximated by

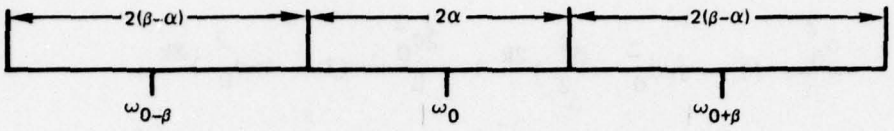

Figure 2. Diagram of a spectrum partition.

$$S(\omega) \sim \sigma_0^2 \frac{\beta^2 + (\omega-\omega_0)^2}{\alpha^2 + (\omega-\omega_0)^2} \qquad \beta << 1 . \tag{74}$$

Likewise, $|F(\omega)|^2$ is approximately

$$|F(\omega)|^2 \sim c^2 e^{\beta} \frac{1}{\beta^2 + (\omega-\omega_0)^2} \qquad \beta << 1 . \tag{75}$$

The time constants for two of the regions are the same so that

$$S(\omega_0) \sim \sigma_0^2 \frac{\beta^2}{\alpha^2} \qquad S(\omega_0 \pm \beta) \sim \sigma_0^2 \frac{2\beta^2}{\alpha^2 + \beta^2} \tag{76}$$

$$(\Delta\omega)_1 |F(\omega_0)|^2 \sim 2\alpha \frac{c^2 e^{\beta}}{\beta^2} \qquad 2(\Delta\omega)_2 |F(\omega_0 \pm \beta)|^2 = 4(\beta-\alpha) \frac{c^2 e^{\beta}}{2\beta^2} .$$

To simplify the expressions, we will assume that $\alpha << \beta$ (this is the case except at extremely low signal-to-noise ratios). The substitution of (76) into (53) and (54) then yields a sum of terms of the form (55). They are, up to the factor $2c^2 e^{-\beta}$,

$$\underline{\text{filter}}: \quad \frac{\alpha}{\beta^2} (1 - 2\mu\sigma_0^2 \frac{\beta^2}{\alpha^2})^{2k} + \frac{1}{\beta} (1 - 4\mu\sigma_0^2)^{2k} \tag{77}$$

$$\underline{\text{output}}: \quad \frac{\sigma_0^2}{\alpha}\left(1 - 2\mu\sigma_0^2\,\frac{\beta^2}{\alpha^2}\right)^{2k} + \frac{2\sigma_0^2}{\beta}\left(1 - 4\mu\sigma_0^2\right)^{2k} . \tag{78}$$

In the above approximation, the filter exhibits two time constants:

$$\tau_1 = \frac{1}{4\mu\sigma_0^2}\,\frac{\alpha^2}{\beta^2}$$

and

$$\tau_2 = \frac{1}{4\mu\sigma_0^2}\,\frac{1}{2} . \tag{79}$$

In contrast, since the slower second term in (78) is negligible (for $\alpha << \beta$), the convergence of the output will depend essentially on only one time constant, $\tau_1 = \frac{1}{4\mu\sigma_0^2}\,\frac{\alpha^2}{\beta^2}$. Note that a closer approximation (more modes) to equation (54), yields a second time constant at $\omega_0 \pm \alpha$:

$\tau_2' = \frac{1}{4\mu}\left(S(\omega_0 \pm \alpha)\right)^{-1} \sim \frac{1}{4\mu\sigma_0^2}\left(\frac{2\alpha^2}{\beta^2}\right)$. Thus, one would expect the logarithmic slope of the actual learning curve for the output to vary between τ_1 and $2\tau_1$ during the period of significant convergence.

The above results offer the following interpretation. Initially, the transfer function of the filter will be narrowly centered about ω_0 (the modes ω for ω close to ω_0 convergence rapidly). With time the filter widens (the slower modes converge). On the other hand, the output converges quickly, since for a narrowband signal, there is little error reduction as

the filter widens. This is particularly true at high SNR ($\beta >> \alpha$); the filter can afford to get quite wide since there is very little noise to let through. At low SNR the optimal filter is narrow ($\beta \sim \alpha$), and this "filter growth" will not usually be observed.

The same considerations hold for large β ($\beta \geq 1$) although the approximation (74) is not valid and must be replaced by (73). For $\beta > 1$, $|F^2(\omega)|$ is approximately constant (equation 72)), all modes are equally important, and convergence will be limited by the larger time constants. On the other hand, the modes of the output learning curve fall off as $S(\omega)$(cf equation (55)), and only those corresponding to $|\omega - \omega_0| \leq 2$ will be significant.

A computer simulation of the ALE was run using pseudo-random noise input with a correlation matrix given by equation (59) and parameters, $L = 64$, $\omega_0 = 1.5$, $\alpha = .2$, $\mu = .0008$, and $\beta = .915$. The results were then ensemble-averaged to obtain $\tilde{W}$. It follows from (63) that SNR = $\frac{\sigma_s^2}{\sigma_0^2}$ = 3.3 dB. A plot of equation (56), time constant versus frequency, appears in Figure 3. It is seen that the fastest mode is about $\tau = 30$ and the slowest, $\tau = 525$. The conditioning number is thus 17.5.

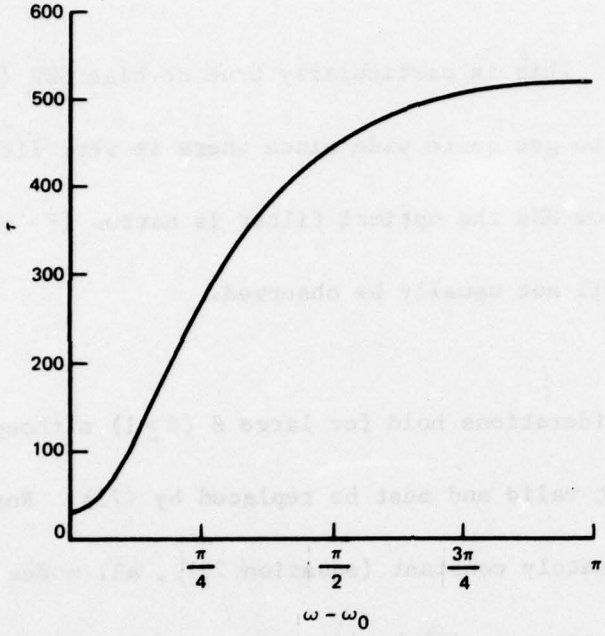

Figure 3. Plot of time constants of the various modes of convergence for the weight vector (equation (56)): $\alpha = 0.2, \beta = 0.915, \mu = 0.008, \sigma_0^2 = 0.5$.

The actual learning curves for the simulation appear in Figures 4 and 5. They are compared with those computed from equations (53) and (54). (It was found that five modes provided a reasonable approximation to the integrals, and that anything over ten modes was virtually indistinguishable from the plotted curves.) The output curves in Figure 5 are in excellent agreement. There is a very slight discrepancy in the weight vector learning curves of Figure 4, which may be attributed to the initial stages of convergence. The independence assumption on which equation (1) is based is not valid for small k. The ultimate effect is the same as if the initial state of the ALE, $\omega(0)$, were not quite 0. In contrast, Figure 5 was constructed by matching conditions at k = 50.

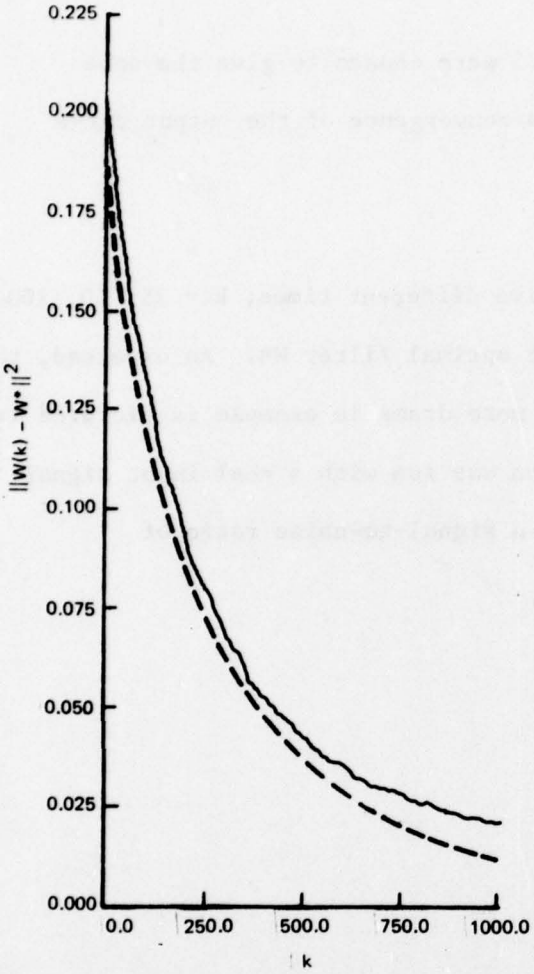

Figure 4. Learning curves for the weight vector. Solid Line = computer simulation of ALE, averaged over a 20-point ensemble. Dotted Line = theoretical curve computed from equation (53).

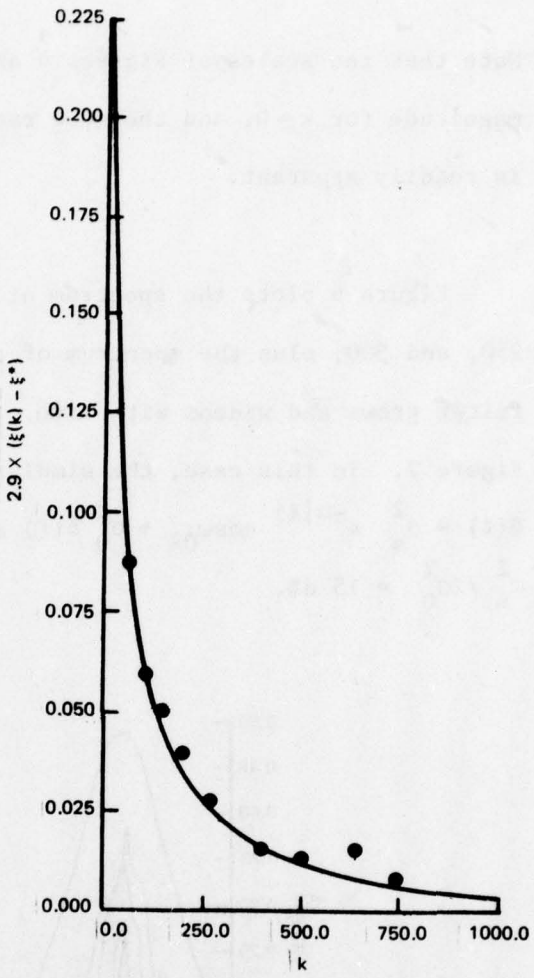

Figure 5. Learning curves for the output. Solid Line = theoretical computed from equation (54). Dots = computer simulation of ALE, averaged over a 20-point ensemble.

Since $\beta = .915$ is rather large, one would expect the overall convergence of the weight vector to depend on the slowest mode. From Figure 3, this is about 500, which corresponds quite well with the simulation in Figure 4. According to theory, the output should converge faster, depending mostly on those modes for which $|\omega - \omega_0| \leq \alpha = .2$. From Figure 3, this value is $\tau \sim 55$ which agrees with the e^{-1} downpoint of Figure 5.

Note that the scales of Figures 4 and 5 were chosen to give the same magnitude for k = 0, and the more rapid convergence of the output curve is readily apparent.

Figure 6 plots the spectrum at five different times; k = 25, 50, 100, 250, and 500; plus the spectrum of the optimal filter W*. As expected, the filter grows and widens with time. A more dramatic example is pictured in Figure 7. In this case, the simulation was run with a real input signal $\phi(\ell) = \sigma_s^2 \; e^{-\alpha|\ell|} \cos\omega_0\ell + \sigma_0^2 \; \Delta(\ell)$ at a signal-to-noise ratio of $\sigma_s^2 / 2\sigma_0^2 = 15$ dB.

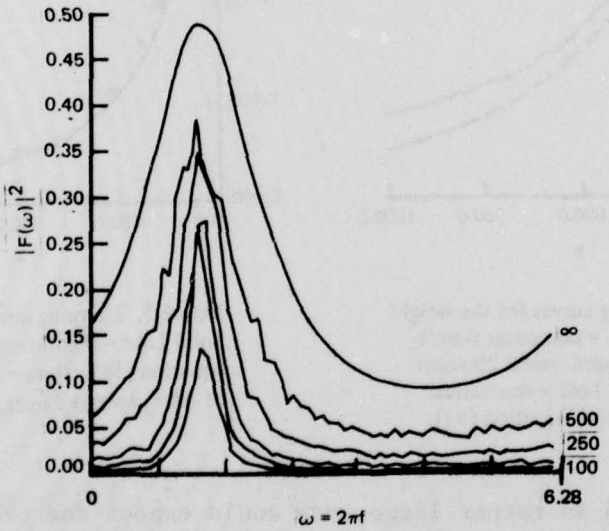

Figure 6. Power spectrum of $\tilde{W}(k)$ for five different times (k = 25, 50, 100, 250, and 500), and the optimal weight vector $W^*(k=\infty)$. The input correlation function of the ALE was given by equation (59) with $\alpha = 0.2$, $\beta = 0.915$, L = 64, $\omega_0 = 1.5$, and $\mu = 0.0008$. The results were averaged over a 100-point ensemble.

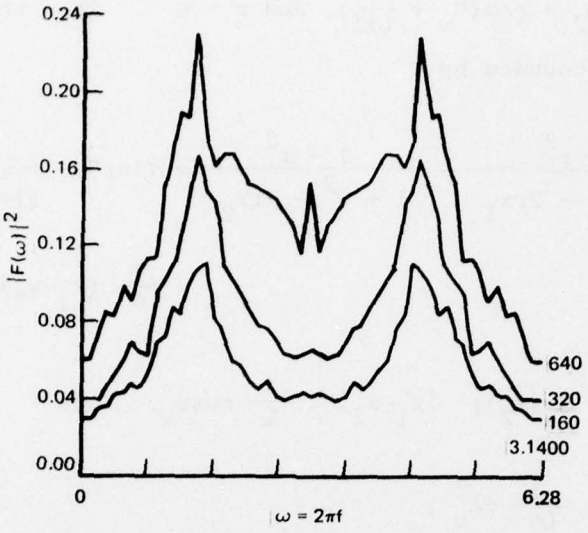

Figure 7. Power spectrum of W(k) at three different times (k=160, 320, and 640). The input correlation function was $\sigma_s^2 e^{-\alpha|\ell|^1} \cos \omega_o \ell + \sigma_0^2$ with $\alpha = 0.2, \beta = 2.01$, L = 64, $\omega_o = 1.5$, and $\mu = 0.0016$. The results were averaged over a 200-point ensemble.

Finally, it should be recalled that the theory of the previous section is an approximation for large L (L = 64 in the above simulation). For the case of a signal pole spectrum, we can explicitly derive the asymptotic dependence of the eigenvalues on L. From [5], the eigenvalues of R_s (matrix (59) with $\sigma_0 = 0$) satisfy

$$\alpha \leq \lambda_1 \leq S(\frac{\pi}{L+1}) \leq \lambda_2 \leq S(\frac{2\pi}{L+1}) \leq \quad . \quad . \quad . \tag{80}$$

Let $x_1 = \cos\theta_\nu$, $x_2 = \cos(\theta_\nu + \frac{\pi}{L+1})$, and $r = e^{-\alpha}$. Then the error in estimating λ_ν is bounded by

$$\frac{1 - r^2}{1 + r^2 - 2rx_1} - \frac{1 - r^2}{1 + r^2 - 2rx_2} = (1-r^2)\,\frac{2r(x_1-x_2)}{(1+r^2-2rx_1)(1+r^2-2rx_2)}$$

$$\sim \frac{1}{\alpha}\,(x_1-x_2)S_s(x_1)S_s(x_2); \quad \alpha < 1 .$$

But $S_s(x_1) = \lambda_\nu \sim S_s(x_2)$; $|x_1-x_2| \sim \frac{\pi}{L}\cos\theta_\nu$. Thus

$$\text{error} < \frac{1}{L\alpha}\,\pi\lambda_\nu^2 . \tag{81}$$

Except for very large eigenvalues, the percentage error will be small when $L\alpha >> 1$; i.e., provided the signal is wider than the ALE resolution. For the other extreme, $L\alpha << 1$, the narrowband signal will appear essentially sinusoid.

VI. CONCLUSION

We have examined the convergence of LMS adaptive filters as reflected by the weight vector and filter output. These curves are a sum of modes with decay constants equal to the eigenvalues of the input correlation matrix. It was shown that, for zero initial conditions, the importance of a particular mode depends on the projection of its eigenvector onto the optimal filter W*. For example, in the case of the adaptive line enhancer with an input of M sinusoids in white noise, only the M-dimensional "signal subspace" is relevant to convergence time.

One and two sinusoids were treated in detail. It was shown that a single real sinusoid (two complex) has two pertinent eigenvalues (M equals 2), and that they are approximately equal. Consequently, convergence is often much quicker than would be expected from an examination of all the eigenvalues. The behavior for two sinusoids (four complex) depends on the separation of their frequencies. If they are relatively close, there can be a large disparity in the eigenvalues even in the signal subspace (equation (37)). The associated eigenvectors were examined, and it was found that there exist circumstances under which the slower mode may dominate the learning curve.

A simple correspondence may be set up, for large filter lengths, between the discrete and continuous cases. Indexed by frequency, the eigenvalues of the correlation matrix correspond to the magnitude of the power spectrum, and the projections of their eigenvectors on W* correspond to the magnitude of the filter transfer function. Obvious though this relationship may be, it provides a powerful means of approximating the LMS learning curves. A rough knowledge of the input spectrum suffices to evaluate the conditioning number and thus set bounds on convergence (equation 57)). A small amount of additional analysis yields approximations to the learning curves (equations (53) and (54)).

As an example, we treated the case of a single pole input spectrum. Simple expressions relating bandwidth, SNR, and approximate convergence times were derived (equation (79)). In general, the convergence of the output is faster than that of the weight vector. This is particularly noticeable at high signal-to-noise ratios. It was also observed that the filter is initially narrow and widens with time (during convergence). The more exact expressions for the learning curves, equations (53) and (54) were then evaluated numerically and proved in excellent agreement with a computer simulation.

The techniques which we have developed often alleviate the need for finding the eigenvalues of large matrices in order to analyze the learning curves. In addition, they provide a physical context, the signal spectrum, in which to evaluate convergence. Thus, it is possible, with varying degrees of accuracy dependent on knowledge of the input spectrum, to predict the convergence behavior of the system in question.

REFERENCES

1. B. Widrow, J. McCool, M. Larimore, C. Richardson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter", Proc. IEEE, Vol. 64, pp. 1151-1162, Aug. 1976.

2. J. R. Treichler, "Transient and Convergent Behavior of the Adaptive Line Enhancer", IEEE Trans. Acous., Speech, and Sig. Proc., submitted for publication.

3. J. Zeidler, E. Satorius, D. Chabries, H. Wexler, "Adaptive Enhancement of Multiple Sinusoids in Uncorrelated Noise", IEEE Trans. Acous., Speech, and Sig. Proc., to be published June 1978.

4. B. Widrow, J. Glover, J. McCool et al., "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE, Vol. 63, pp. 1692-1716, Dec. 1975.

5. V. Grenander and G. Szego, Toeplitz Forms and Their Applications, Univ. of Calif. Press, Berkeley, 1958.

6. A. Oppenheim and R. Schafer, Digital Signal Processing, Prentice Hall, Englewood Cliff, N.J., 1975.

7. A. Householder, The Theory of Matrices in Numerical Analysis, Chapter 2, Dover Publications, N.Y., 1964.

8. A. Taylor, Introduction to Functional Analysis, especially Chapter 6, John Wilers and Sons, N.Y., 1958.

9. S. Alexander, E. Satorius, J. Zeidler, "Linear Prediction and Maximum Entropy Spectral Analysis of Finite Bandwidth Signals in Noise", 1978 IEE International Conference on Acoustics, Speech, and Signal Processing, Tulsa, Oklahoma, April, 1978.

10. W. Davenport and W. Root, An Introduction to Random Signals and Noise, Chapter 11, McGraw Hill, N.Y., 1958.

11. R. Hearn, M. Shensa, M. Whaley, "Experimental Measurement of Adaptive Noise Canceller Performance", NOSC Report, Code 632, June 1978.

APPENDIX A

We are concerned with the eigenvectors of the submatrix

$$\tilde{B} = \frac{1}{2} \begin{pmatrix} L & Z \\ \bar{Z} & L \end{pmatrix} \tag{A-1}$$

which has eigenvalues

$$\lambda_{\pm} = \frac{L}{2} \pm \frac{|Z|}{2} \tag{A-2}$$

where

$$Z = \frac{1 - e^{i\Delta L}}{1 - e^{i\Delta}} . \tag{A-3}$$

Up to order Δ^2,

$$Z \sim L + \frac{iL^2\Delta}{2} - \frac{L^3\Delta^2}{6} \tag{A-4}$$

$$\lambda_+ \sim L$$

$$\lambda_- \sim 0 .$$

Let $\begin{pmatrix} 1 \\ b \end{pmatrix}$ be eigenvectors of $\tilde{B}$, then

$$\tilde{B} \begin{pmatrix} 1 \\ b_{\pm} \end{pmatrix} = \lambda_{\pm} \begin{pmatrix} 1 \\ b_{\pm} \end{pmatrix}$$

which implies

$$L + bZ = 2\lambda \ , \tag{A-5}$$

and, hence,

$$b_+ \sim 1 - \frac{i\Delta L}{2} \tag{A-6}$$

$$b_- \sim -1 + \frac{i\Delta L}{2} \ .$$

We now calculate W*. From equation (27), with $\sigma_1^2 = \sigma_2^2$, $\delta = 1$, and $\omega_1 \sim \omega_2$, P in the current basis is proportional to $\begin{pmatrix}1\\1\end{pmatrix}$. Also, R is given by

$$R = \sigma_0^2 + \sigma_1^2 \tilde{B} = \sigma_0^2 \begin{pmatrix} 1+Lq & qZ \\ q\bar{Z} & 1+Lq \end{pmatrix} \tag{A-7}$$

where

$$q = \frac{\sigma_1^2}{2\sigma_0^2} \ .$$

Finally, $W^* = R^{-1}P$ is proportional to

$$\begin{pmatrix} 1+Lq & -qZ \\ -q\bar{Z} & 1+Lq \end{pmatrix} \quad \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix} 1+Lq-qZ \\ 1+Lq-q\bar{Z} \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 - \frac{iq\Delta L^2}{2} + \frac{L^3\Delta^2 q}{6} \\ 1 + \frac{iq\Delta L^2}{2} + \frac{L^3\Delta^2 q}{6} \end{pmatrix} \ . \tag{A-8}$$

Define

$$E^{+} = \begin{pmatrix} 1 \\ b_{+} \end{pmatrix} \qquad E^{-} = \begin{pmatrix} 1 \\ B_{-} \end{pmatrix} \quad , \tag{A-9}$$

so that $||E^{+}|| = ||E^{-}||$. Then, up to a constant factor

$$|E^{+} \cdot W^{*}|^{2} \sim (2 + \frac{L^{3}\Delta^{2}q}{3})^{2} \tag{A-10}$$

$$|E^{-} \cdot W^{*}|^{2} \sim q^{2}\Delta^{2}L^{4}$$

where $\Delta L << 1$.

APPENDIX B

For mathematical background and material related to the subject matter in this appendix, the reader is referred to references [5] - [8]. We proceed to prove relation (46).

Define the strong matrix norm by

$$||A|| = \sup_{x} \frac{||Ax||}{||x||} . \tag{B-1}$$

It follows from this definition [7] that the norm of a finite dimensional Hermitian matrix B is given by

$$||B|| = \lambda_{max} \tag{B-2}$$

$$= \text{maximum eigenvalue of } B .$$

Let y be a vector in ℓ^2, i.e., $\sum_{\ell=0}^{\infty} y_\ell^2 < \infty$, then there exists $Y(\omega)$, the inverse transform of y, such that

$$y_\ell = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y(\omega) e^{i\omega\ell} d\omega . \tag{B-3}$$

The convolution theorem for Fourier series and Plancherel's theorem imply

$$\left|\left| \sum_{k=0}^{\infty} \phi_{\ell k} y_k \right|\right|^2 = \sum_{\ell=0}^{\infty} (\phi_{\ell k} y_k)^2$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} |Y(\omega)|^2 \, |S(\omega)|^2 \, d\omega \ . \qquad \text{(B-4)}$$

Thus

$$||\phi|| = \sup_{y} ||\phi y|| / ||y||$$

$$= \max_{\omega} |S(\omega)| \ . \qquad \text{(B-5)}$$

Let R^L be the operator ϕ restricted to the L-dimensional subspace of the first L basis vectors:

$$(R^L y)_\ell = \begin{cases} \sum_{k=0}^{L-1} \phi_{\ell k} \, y_k & \ell \leq L-1 \\ 0 & \ell \geq L \ . \end{cases} \qquad \text{(B-6)}$$

Thus, R^L is identical to R of equation (44). If we define the projection operator Q^L by

$$(Q^L y)_\ell = \begin{cases} y_\ell & \ell \leq L-1 \\ 0 & \ell \geq L \end{cases} \qquad \text{(B-7)}$$

then

$$R^L = Q^L \, \phi \, Q^L \ .$$

Thus,

$$||R^L|| \leq ||Q^L||\ \ ||\phi||\ ||Q^L||$$

$$= ||\phi|| \ . \qquad \text{(B-8)}$$

Similar expressions hold for the infinum of R^L and ϕ:

$$\lambda^L_{\min} = \min_y \frac{||R^L y||}{||y||}$$

$$= \min_y \frac{||Q^L \phi Q^L y||}{||y||}$$

$$\geq \inf_y \frac{||\phi Q^L y||}{||Q^L y||}$$

$$\geq \inf_z \frac{||\phi z||}{||z||}$$

$$= \min_\omega |S(\omega)| \qquad \text{(from (B-4))} \qquad \text{(B-9)}$$

It was assumed that $G(z)$ has no zeroes on the unit circle; thus, we may write $\min_\omega |S(\omega)| = a > 0$. Then

$$||(R^L)^{-1}|| = \text{max eigenvalue } (R^L)^{-1}$$

$$= \frac{1}{\lambda^L_{\min}}$$

$$\leq \frac{1}{a} \ . \qquad \text{(B-10)}$$

Also, it follows from (B-5) and (B-9) (cf [8] that ϕ^{-1} exists and is bounded.

Finally, we note that R^L converges weakly to ϕ; i.e., that

$$\lim_{L\to\infty} ||(R^L-\phi)y|| = 0 \qquad \forall y \ . \tag{B-11}$$

This follows from

$$\begin{aligned} ||(R^L-\phi)y|| &= ||Q^L\phi Q^L y-\phi y|| \\ &= ||Q^L(\phi Q^L-\phi)y + (Q^L\phi-\phi)y|| \\ &\le ||\phi(Q^L-I)y|| + ||(Q^L-I)\phi y|| \\ &\le ||\phi||\ ||(Q^L-I)y|| + ||(Q^L-I)\phi y|| \end{aligned} \tag{B-12}$$

and the fact that

$$||Q^L-I)y||^2 = \sum_{\ell=L}^{\infty} y_\ell^2$$

converges to zero as $L \to \infty$.

Let W^L be the solution of the L-dimensional Wiener-Hopf equation (5)

$$W^L = (R^L)^{-1}Q^L P \tag{B-13}$$

where P is defined by (42). Let $f = \phi^{-1}P$ and $P^L = Q^L P$. Then

$$||f-W^L|| = ||\phi^{-1}P - (R^L)^{-1}P^L||$$

$$\leq ||(\phi^{-1} - (R^L)^{-1})P|| + ||(R^L)^{-1}(P-P^L)|| .$$

Also,

$$||(\phi^{-1} - (R^L)^{-1})P|| = ||(R^L)^{-1}(R^L-\phi)\phi^{-1}P||$$

$$\leq ||(R^L)^{-1}|| \; ||R^L-\phi)f|| .$$

Hence,

$$||f-W^L|| \leq \frac{1}{a} \left(||(R^L-\phi)f|| + ||P-P^L||\right),$$

and (B-12) implies

$$\lim_{L\to\infty} ||f-W^L|| = 0 . \qquad \text{(B-14)}$$

We now prove (53) and (54). Let

$$\hat{A} = I-2\mu\phi$$

$$A = I-2\mu R^L . \qquad \text{(B-15)}$$

Then the substitution of $\hat{A}$ and A for ϕ and R^L plus an induction on the inequality (B-12) implies that A^k converges weakly to $\hat{A}^k$ and thus

$$\lim_{L\to\infty} ||A^k y|| = ||\hat{A}^k y|| . \qquad \text{(B-16)}$$

Also,

$$||A^k W^L|| \leq ||A^k W^L - A^k f|| + ||A^k f - \hat{A}^k f|| + ||\hat{A}^k f|| .$$

Hence,

$$\lim_{L\to\infty} ||A^k W^L||^2 = ||\hat{A}^k f||^2 . \tag{B-17}$$

But the convolution theorem implies

$$(\hat{A}^k)_{\ell m} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1-2\mu S(\omega)\right)^k e^{i\omega(\ell-m)}\, d\omega \tag{B-18}$$

and Plancherel's theorem gives

$$||\hat{A}^k f||^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1-2\mu S(\omega)\right)^{2k} |F(\omega)|^2 d\omega \tag{B-19}$$

with $F(\omega)$ defined as in (52). Substitution of (8) and (B-19) into (B-17) yields equation (53). A similar calculation using (8) and (11) gives equation (54).